



## VIII Congreso Latinoamericano de Sociedades de Estadística

Octubre de 2008

### Sesionador Web dirigido al estudio de sitios web culturales: Diseño e Implementación del paquete RWeb Sessionizer

Esther Hochsztain, Andrómaca Tasistro y Carolina Asuaga

#### Resumen ampliado

Resulta imprescindible para las organizaciones evaluar que tan efectivo es su sitio web. Usualmente se analiza solamente los datos referidos al número de páginas consultadas y la cantidad de visitantes en el sitio web, así como en qué partes del sitio web el usuario clickea cuando elige incluir un producto en el carro de compras, comprar un ítem, u otras acciones de compra, eventos de registro, vista de productos, acciones de pago, entre otras. Sin embargo, lo anterior brinda una visión muy básica y se necesitan aplicaciones más avanzadas para estudiar características adicionales.

En este artículo, se muestra que un sesionador web (Web Sessioner) resulta de fundamental importancia en Web Usage Mining. Consiste en la aplicación de técnicas de Data Mining para la identificación de patrones de uso de un sitio web. Se describe una sesión de usuario como aquella formada por un conjunto de objetos consultados por un mismo usuario durante una misma visita a un sitio web.

Entre las diversas metodologías existente para la identificación de usuarios, en esta propuesta se usa la pareja dirección IP y el agente desde donde se realiza la solicitud. Este procedimiento presenta como ventaja que siempre está disponible y no se necesita tecnología adicional. Una vez identificado

se genera una lista de sesiones utilizando heurísticas orientadas al tiempo o a la navegación, luego de descartar de crawlers. Se utilizó el paquete RWeb Sessionizer donde se implementó en R la metodología propuesta inicialmente por R. Cooley, y que ha tenido varios refinamientos posteriores. El sessionizer implementado posee las siguientes características: timeout parametrizable, elimina software robots (crawlers), asigna un identificador a cada sesión, genera datos de resumen del proceso de sesionalización.

El caso de estudio fue el Teatro Solís, principal referente de la cultura montevideana. La elección de la organización no es casual, ya que el Teatro Solís es también una institución de fuerte contenido

simbólico y forma parte del patrimonio cultural de los uruguayos en general y los montevideanos en particular

En los resultados preliminares se obtuvo: Cantidad de das: 3: Cantidad de líneas: 115744T; Timeout: 10 minutos Cantidad de sesiones (eliminados crawlers): 5248, promedio de bytes por sesión: 8722 Duración promedio de la sesión: 4,4 minutos.

El artículo concluye que contar con este paquete es un primer paso para comprobar si el sitio web del Teatro Solís está cumpliendo con los objetivos previstos. Con los resultados preliminares se pudo apreciar que la página principal está cumpliendo el objetivo de brindar los datos necesarios para la mayoría de los usuarios.

El Teatro Sols es patrimonio de todos los uruguayos. Una correcta gestión de su sitio web, en el que además de promover y difundir los diversos espectáculos artísticos, se propicie la cohesión social, así como la generación y reafirmación de valores simbólicos compartidos, es un tema no menor tanto en los objetivos del Teatro como en el beneficio de la sociedad toda.

Como trabajos futuros se propone incorporar el módulo de análisis de sesiones en base a datos de resumen y una galera de gráficos; y establecer tanto un método de consulta interactivo como una interfaz de usuario gráfica que facilite el uso del paquete.

**Palabras clave:** sitio web, Sesionador Web, RWeb Sessionizer, clientes, consumidores, organizaciones culturales, teatros

## Referencias

Abuhamdieh, A: Kendall JE, and Kendal KEI. 2007. "E -Commerce Opportunities in the Nonprofit Sector: The Case of New York Theatre Group." *International Journal of Cases on Electronic Commerce*, Vol.3, Issue 1.

Andrejko, Anton, Michal Barla, M´aria Bielikov´a, and Michal Tvarozek. 2007. "User Characteristics Acquisition from Logs with Semantics." ISIM.

Asuaga, C., Cambeiro, P., & Cami, M. 2007. Gestión de Teatros Públicos *Quantum, Revista de Contabilidad, Economía y Administración*, 2(1).

Asuaga, C., & Rausell-Köster, P. 2006. Management of Cultural Organization: *Revista iberoamericana de contabilidad de gestión*, 4(8), 83-104.

Budiño, G. 2004. *Sistemas de información para la satisfacción de clientes*. Universitario Autónomo del Sur-Uruguay

Castellano, G., A. M. Fanelli, and M. A. Torsello. 2007. "LODAP: a log data preprocessor for mining web browsing patterns." AIKED'07: Proceedings of the 6th Conference on 6th WSEAS Int. Conf. on Artificial

Intelligence, Knowledge Engineering and Data Bases. Stevens Point, Wisconsin, USA: World Scientific and Engineering Academy and Society (WSEAS), 12–17.

Cowen, Tyler. 2008. "Why everything has changed: the recent revolution in cultural economics." *Journal of Cultural Economics*. Publisher Springer Netherlands 32 (4): 261–273 (diciembre).

Goldfarb, Avi, and Jeff Prince. 2008. "Internet adoption and usage patterns are different: Implications for the digital divide." *Information Economics and Policy* 20 (1): 2–15 (March).

Hall, John C. 2007. "The Old Operating Theatre, Museum and Herb Garret website." *Journal of Surgery*. Vol. 77 Issue 12.

Hartley, John. 2008. "The future is an open future. cultural studies at the end of the 'long twentieth century' and the beginning of the 'chinese century'." CCI/ Feast Joint Research

Hochsztain, E., Ramirez, R., Tasistro, A., & Asuaga, C (2008). Aplicaciones de web mining al análisis del comportamiento de los usuarios del sitio web del Teatro Solís.

Hochsztain, E. y Asuaga C (2008): Cultura e Internet: Actividad de extensión vinculada al análisis del sitio web del Teatro Solís. *I Jornadas de Extensión del Área Social*. Udelar

Jin-hua Zhu, and Jun-jie Chen. 2008. "Research on Method for Session Identification in Web Log Mining." *Journal of Taiyuan University of Technology*.

Kendall, J. E. 2006. "Theatres, Metaphors, and E-Collaboration: An Examination of Web-Based Cooperation of Regional Nonprofit Theatres." *Theatres. International Journal of e-Collaboration*, Vol. 2, Issue 1.

Komito, Lee. 2008. Chapter Information Society Policy of Knowledge policy: challenges for the 21st century, edited by Greg Hearn and David Rooney, 83–97. Edward Elgar Publishing Limited.

Lazarinis, Fotis, Dimitris Kanellopoulos, and Petros Lalos. 2008. "Heuristically Evaluating Greek e-Tourism and e-Museum Websites." *Electronic Journal Information Systems Evaluation* Volume 2008.

Liu, Haibin, and Vlado Keelj. 2007. "Combined mining of Web server logs and web contents for classifying user navigation patterns and predicting users future requests." *Information Processing Management* Volume 61, Issue 2.

Marty, P. 2007. "Museum Websites and Museum Visitors: Before and After the Museum Visit." *Museum Management and Curatorship*, Volume 22, Issue 4.

Mason, D, and C McCarthy. 2008. "Museums and the culture of new media: an empirical model of New Zealand museum websites." *Museum Management and Curatorship*, Volume 23, Issue 1.

Menasalvas, E., Millán, S., Pérez, M., Hochsztain, E., Robles, V., Marbán, O., Peña, J. and Tasistro, A., 2003. Beyond user clicks: an algorithm and an agent-based architecture to discover user behavior. In *First European Web Mining Forum* (p. 80).

Menasalvas, E., Millán, S., & Hochsztain, E. 2002. A granular approach for analyzing the degree of afability of a web site. In *International Conference on Rough Sets and Current Trends in Computing RSCTC2002*.

Peacock, Alan, and Ilde Rizzo. 2008. *The Heritage Game. Economics, Policy, and Practice*. Edited by Oxford University Press. Oxford University Press.

Rausell, Pau. 2002. "Los mercados culturales y el desarrollo de la nueva economía." *Debate Culturales*. Universitat Oberta de Catalunya.

Slack, F., J. Rowley, and S. Coles. 2008. "Consumer behaviour in multichannel contexts: the case of a theatre festival." *Internet Research*. Volume 18 Issue:1.

Veltman, Kim H. 2005. "Desafíos en las aplicaciones de las TIC/TCU en el patrimonio cultural." *Digithum: Las humanidades en la era digital*, vol. 7.

Wang, Yiwen, Lora M. Aroyo, Natalia Stash, and Lloyd Rutledge. 2007. "Interactive User Modeling for Personalized Access to Museum Collections: The Rijksmuseum Case." *Study Lecture Notes in Computer Science*. Volume 4511.